Measure and Mitigate the Dimensional Bias in Online Reviews and Ratings

Research-in-Progress

Yong Ge
University of North Carolina at Charlotte
9201 University City Blvd, Charlotte, NC 28223
yong.ge@uncc.edu

Jingjing Li
University of Virginia
125 Ruppel Drive Charlottesville, VA 22903
Jingjing.li@comm.virginia.edu

Abstract

Online word-of-mouth in the form of online reviews and ratings is an increasingly important resource for consumers to acquire product information for their purchase decision. However, dimensional review bias, originated from consumer heterogeneity and their multidimensional product preferences and experiences, have been shown to undermine the information transfer among consumers. Through a novel text mining approach, we identify and quantify two types of dimensional biases from textual reviews: dimensional preference bias and dimensional rating bias. We also introduce a quantitative method to mitigate the dimensional rating bias. We examined the effectiveness and applicability of our bias measures and de-bias method in the context of multi-dimensional and single-dimensional rating systems. Specifically, we focused on the hotel reviews from TripAdvisor.com and Expedia.com. Our preliminary results show promising theoretical and managerial contributions.

Keywords: dimensional bias, online review, review bias, rating bias

Introduction

Online word-of-mouth (WOM) in the form of reviews and ratings has been an increasingly important resource for consumers to acquire product information and to support their purchase decisions. For instance, a recent survey conducted by Dimensional Research revealed "90% of consumer buying decisions are influenced by online reviews" (Dimensional Research 2013). There are two major types of online review systems: single-dimensional rating (SR) system and multi-dimensional rating (MR) system. A review of an SR system includes an overall numerical rating and a textual review. Examples of SR systems include Amazon.com, Expedia.com, and Walmart.com. On the other hand, a review of an MR system usually consists of multiple numerical ratings on various product dimensions and a textual review. TripAdvisor.com and BeerAdvocate.com are common examples of such a system. Based on numerical ratings and textual reviews, several review measures have been proposed to empirically test the economic value of online reviews. Many studies have focused on creating measures from numerical ratings, e.g., valence, volume and dispersion (Dellarocas et al. 2007, Duan et al. 2008a, Duan et al. 2008b, Liu 2006, Godes and Mayzlin 2004). Recently, a few studies tried to extract measures from textual reviews, e.g., readability, subjectivity and spelling error (Ghose et al. 2011).

Beyond the review measures mentioned above, review bias has attracted more and more attention from researchers and practitioners. Two types of biases have been revealed in the previous literature. The first one is *social influence bias*, or the herding effect (Krishnan et al. 2014, Muchnik et al. 2013 and Wang et al. 2014). The social influence bias describes a situation that a consumer changes his/her intended product evaluation after seeing ratings from others. For example, in Krishnan et al. (2014)'s experiment, the participants' ratings are significantly changed after seeing the historical median ratings. The second type of bias, which is less studied, is *dimensional bias* (Liu et al. 2014). The dimensional bias occurs if a consumer only emphasizes his/her opinions on a few dimensions of a product, or a consumer's overall rating is driven by some extremely positive or negative experiences of a few product dimensions. For instance, Liu et al. (2014) found that a consumer's overall rating of a restaurant is skewed towards the least satisfactory restaurant dimension, particularly in an SR system. Past studies have shown that both social influence bias and dimensional bias could significantly undermine information transfer among consumers (Muchnik et al. 2013, Liu et al. 2014, Wang et al. 2014).

While much research efforts have been dedicated to understanding social influence bias (Krishnan et al. 2014, Muchnik et al. 2013 and Wang et al. 2014), little has been done on the dimensional bias. After reviewing related psychology literature, we find that the dimensional bias is potentially coming from two sources. The first is the heterogeneous multi-dimensional product preferences (Ghose et al. 2012), indicating that consumers are more likely to express their opinions on product dimensions they care about. The second one is the hemeostase utility (Hennig - Thurau et al. 2004) of writing a review, which means that consumers are likely to focus more on extreme negative or positive experience when writing reviews due to a need to restore balance in their lives. In light of these findings, we identify two types of dimensional bias: *dimensional preference bias* and *dimensional rating bias*. The dimensional preference bias refers to a consumer's tendency to express opinions on one or a few product dimensions when writing a review. The dimensional rating bias refers to the situation that a consumer's overall rating is skewed towards one or few dimensions experiencing extreme sentiment and ratings. Studying both types of dimensional biases requires obtaining information from textual reviews, which is especially important for SR systems. To the best of our knowledge, no studies have focused on identifying dimensional bias from larege-scale textual reviews using an automated method.

In this paper, we aim at automatically extracting and quantifying dimensional preference and rating bias for both SR and MR systems by mining textual reviews. Additionally, we propose a de-bias method to mitigate the dimensional rating bias. Our solution consists of two steps. In the first step, we use text mining to identify dimensional product mentions in the textual review and a customized sentimental analysis to estimate the rating of each dimension. In the second step, two quantitative bias measures along with a de-bias method are developed based on the information from step one. In order to demonstrate the applicability and effectiveness of our method, we use two hotel review datasets collected from Expedia.com and TripAdvisor, which are corresponding to SR and MR system respectively. Preliminary results reveal both types of dimensional bias in SR and MR systems. The dimensional bias in SR systems is significantly larger than that in MR. In addition, by segmenting the dimensional bias in reviews with different overall ratings, we find that both types of dimensional bias become more significant in reviews with extreme overall ratings (e.g., 1 in a five-star rating system).

Literature Review and Solution Development

Dimensional Preference Bias

Given the multi-dimensional characteristics of products, consumers often exert heterogeneous preferences over product dimensions, leading to a *dimensional preference bias* in product reviews. Writing online reviews serves as a channel to express the post-purchase satisfaction (Hennig-Thurau et al. 2004). According to the Expectation-Confirmation Theory (Oliver 1980), this satisfaction is determined by the degree to which the perceived performance meets, exceeds or falls below one's prior expectation, i.e. the disconfirmation judgment. Due to the heterogeneous preferences on product dimensions, consumers will assign varying weights to different dimensional disconfirmation judgments. This difference would reflect in the disproportional dimensional mentions in textual reviews. In a hotel review example, for a customer who cares deeply about service quality, he/she is more likely to write about their prior expectation, experience and disconfirmation judgments on service quality. As a result, for review readers who do not stress the service quality that much, they may perceive this entire review as irrelevant, thus discounting the reviewer's overall rating and evaluation.

Dimensional Rating Bias

Another dimensional bias identified in our study, so-called *dimensional rating bias*, is related to a customer's heterogeneous experiences toward different dimensions. According to the Balance Theory (Heider 1946, 1958, Newcomb 1953), people have a basic desire for balance in their lives (e.g., Zajonc 1971). Thus, when experiencing a strong unbalance from either a strong positive or negative consumption experience, consumers need to restore the equilibrium by expressing related positive emotions and negative feelings in reviews. This motive is referred as Homeostase Utility (Hennig-Thurau et al. 2004). Therefore, a consumer's overall satisfaction is likely to skew towards dimensions with extreme sentiment. It is also worth noting that this degree of skewness may be different between positive and negative sentiment. For example, several studies (Anderson and Sullivan 1993; Liu et al. 2014) have shown that negative dimensional ratings exert a stronger influence on overall satisfaction than positive dimensional ratings. As a result, the overall rating of a product will be more skewed toward dimensions with the extremely negative sentiment.

Single-Dimensional Systems and Multi-Dimensional Systems

Due to the prevalence of multi-dimensional product characteristics and dimensional bias, MR systems have been found to be more effective in transferring consumer experience than SR systems (Archak et al. 2011, Godes and Silva 2012, Moe and Schweidel 2012 , Liu et al. 2014). Meanwhile, several IS researchers have proposed new multi-dimensional ranking systems and recommender systems. For instance, Ghose et al. (2012) proposed a ranking system that could reflect a consumer's heterogeneous dimensional preferences for hotel characteristics. Adomavicius et al. (2010, 2005 and 2011) developed a multi-dimensional recommender system with a new type of query language to accommodate a consumer's dimensional preference. Regardless of the increasing popularity of MR systems, SR systems (e.g. Amazon.com and Expedia.com) still occupy a significant proportion of the existing review platforms. Since there is no way to infer dimensional preferences and bias from one overall numerical rating, it is important to develop a method to obtain multi-dimensional information from textual reviews.

Dimension Mining and Rating

Dimension mining is to discover dimensions of an entity (e.g., hotel, restaurant) from a large corpus (e.g., a set of online reviews) and dimension rating is to estimate the possible ratings of user on these dimensions. In the literature, there are some existing works on dimension identification and rating. For example, Moghaddam and Ester (2011, 2012) propose LDA based models to mine the underlying dimensions of product and estimate their corresponding dimension ratings based on the sentiment phrases. Titov and McDonald (2008) try to enhance the coherence between the extracted topics and corresponding dimensions, which basically threw light on the dimension identifications. And Wang et al. (2011) combine LDA and a rating regression approach to automatically uncover the latent aspects and the ratings on each aspect with textual reviews and overall ratings.

Research Gap and Solution Overview

To the best of our knowledge, limited research has focused on extracting and quantifying these bias from textual reviews. For instance, although Liu et al. (2014) proposed a way to compare the dimensional rating bias in both SR and MR systems, they only employ numerical ratings. Because textual reviews often contain richer information (e.g. context) about consumer experiences, we believe that extracting bias from textual reviews will be more valid and informative.

In this study, we achieve this goal by proposing a novel **two-stage** data mining approach (Figure 1). The first stage involves inferring dimensional preferences and calculating related sentiment ratings from textual data. In the second stage, we first introduce two measures to quantify dimensional preference bias and dimensional rating bias. We then propose a de-bias method to mitigate dimensional rating bias. Since our solution uses textual reviews and is independent of user-provided dimensional ratings, it is applicable to both SR and MR systems.

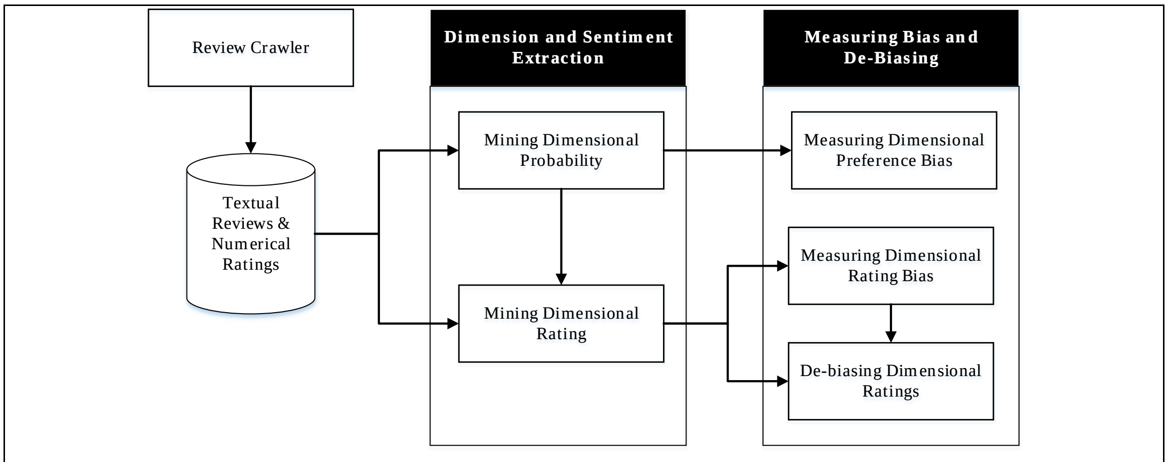

Figure 1. The Overview of Solution

Our approach has several key characteristics that make it quite different from other works on review bias: (a) we identify two types of dimensional bias from textual reviews. As we discussed, textual reviews are often more informative than numerical ratings, and are particularly useful for identifying dimensional bias in SR systems. (b) we are the first study to propose quantitative dimensional bias measures from textual reviews. These measures allow behavior researchers to empiricially test the consequences of dimensional bias. (c) we propose a de-bias method to mitigate the dimensional rating bias. This could be directly leveraged by practitioners to achieve efficient positive WOM transferring.

Solution Details

Dimension and Sentiment Extraction

Mining Dimensional Probability. Before measuring the dimensional preference bias, we need to first identify product dimensions and their relative expressiveness (referred as dimensional probabilities onwards) in the textual review. Although there are some existing works (Moghaddam and Ester 2011, 2012, Titov and McDonald 2008) in the literature, many of these methods work as a black box. Thus, it is difficult to explicitly explain the results and reveal business insights. To this end, we design an expectation-maximization (EM) algorithm to identify dimensional probabilities. The basic idea is that if a review has more sentences including dimension-related keywords, a higher expressiveness (i.e., dimensional probability) will be assigned to this dimension. As the first step of our algorithm, we manually select a keyword set for each product dimension, which is used to find the expressiveness of product dimensions in the textual reviews. As the EM algorithm runs, the keyword set will be automatically expanded. Compared with existing works (Moghaddam and Ester 2011, 2012, Titov and McDonald 2008), it is much easier to explain the generated dimensional probabilities with our method.

Figure 2 shows the details of EM algorithm. We first compute a weighted count of dimension-related keywords for each sentence in a review, where the weight of a keyword is the probability that this keyword belongs to a dimension (the probability of initial seed keyword is set as 1). We then obtain the dimensional probability of a sentence by normalizing the weighted count. Based on the weighted count for each sentence, we calculate the association q between a dimension and a word based on Equation 1:

$$q^k(w, V^k) = \sum_{x=1}^{|V^k|} q_x^k \frac{C}{\max(C_w, C_x)}, \tag{1}$$

where V^k denotes the set of keywords of dimension k; $|V^k|$ denotes the number of keywords in the set; q_x^k is the association of keyword x for dimension k; C denotes the number of sentences including both word w and keyword x; C_w (C_x) represents the number of sentences including word w (x). Subsequently, we include these words that have association values higher than a threshold into dimensional keyword sets.

These steps are repeated until the dimension-related keyword set becomes steady. After the iteration of dimension assignment ends, we will get K-tuple aggregated dimensional probability (Pr^1, Pr^2, ..., Pr^K) for each review. As the aggregation may lead the probability beyond 1, we normalize the dimensional probability for each review as: $P^k = \frac{Pr^k}{\sum_{k=1}^{K} Pr^k}$, where Pr^k denotes the aggregated dimensional probability for one review.

Algorithm: Dimensional Probability Identification Algorithm

Input: A collection of reviews {d_1,...,d_i,...,d_N}, K sets of seed dimension keywords {V^k} (k=1,...,K),

Output: K-tuple dimensional probability for each review

Step 1: Split each review d_i into sentences {s_1,...,s_j,...,s_{Mi}}, where M_i represents the number of sentences in d_i.

Step 2: Match the dimension keywords in each sentence and record the matching hits for each dimension

Step 3: Calculate the weighted count of dimension keywords by aggregating the weights of matched keywords for each pair of dimension k and sentence s_j of a review d_i as $WC_i^{jk} = \sum_{y=1}^{Y} q_y^k$, where q_y^k denotes the weight for one keyword of dimension k and Y is the total number of matched keywords for a pair of dimension k and sentence s_j of a review d_i ,and then obtain the dimensional probability for a sentence by the normalization over all dimensions as $Ps_i^{jk} = \frac{WC_i^{jk}}{\sum_k WC_i^{jk}}$

Step 4: Calculate the association q^k of each word with the keyword set of dimension k based on the Equation 1.

Step 5: Rank the words of each dimension with respect to their q^k values and join these words with q^k values higher than a threshold into their corresponding dimension keyword set for each dimension.

Step 6: If the dimension keyword set is not changed, go to Step 7; otherwise, go to Step 1;

Step 7: Aggregate the dimensional probability over sentences to get the aggregated dimensional probability for each review as $Pr_i^k = \sum_j Ps_i^{jk}$.

Figure 2. An EM Algorithm for Dimensional Probability Identification

Mining Dimensional Rating. After getting the dimensional probabilities from textual review, we continue to find the dimensional rating for these identified dimensions. We achieve this by applying a sentiment analysis model (Hu and Liu 2004) to each dimension. The underlying principle of this model is that a higher rating will be assigned to a dimension if more sentences related to this dimension include positive words, and vice versa. Mathematically, for a sentence s_j from a review d_i, we quantify its sentiment via aggregating the polarity of all words in the sentence as: $Pol_i^j = \sum_x pol_x * f_x$, where pol_x denotes the polarity of a word x in the sentence and f_x is the frequency of word x. In the sentiment analysis literature, researchers have collected different sets of benchmarking polarity of words (NLTK 2015, UIC 2015). In this paper, we use the labeled vocabulary provided by UIC (UIC 2015), which consist of a set of words labeled with a polarity value. The polarity value is between 0 and 1, and a bigger value means more positive sentiment. As we have a normalized dimensional probability Ps_i^{jk} for each sentence in a review as shown in the EM algorithm, we define the following equation to aggregate the polarity over all sentences in a review to get the dimensional rating: $Pol_i^k = \sum_j^{M_i} Pol_i^j * Ps_i^{jk}$, where M_i still denotes the number of sentences in review d_i. As we expect that the dimension rating ranges from 0 to 5, we will map Pol_i^k to this range via a linear function. After such mapping, Pol_i^k is considered as the estimated rating of dimension k for a review d_i.

Measuring Bias and De-biasing

Measuring Dimensional Preference Bias. After identifying a K-tuple $\{P^k\}$ (k =1...K) of dimensional probability for each review, we can measure the dimensional preference bias. A review with high bias tends to have a probability distribution skewing to a few dimensions. Based on this intuition, we define the following measurements to quantify the *dimensional preference bias*:

$$B_t^1 = \{\max\{P^k\} - mean(\{P^k\} \vdash \max\{P^k\})\} \Big/ mean(\{P^k\} \vdash \max\{P^k\}) \text{ or}$$

$$B_t^2 = \{\text{mean}(\max\_2\{P^k\}) - mean(\{P^k\} \vdash \max\_2\{P^k\})\} \Big/ mean(\{P^k\} \vdash \max\_2\{P^k\}), \quad (2)$$

where $\max\{P^k\}$ denotes the maximum value in K-tuple and $\max\_2\{P^k\}$ denotes the first two maximum values in K-tuple. The symbol $\vdash$ indicates taking elements from a set, thus $\{P^k\} \vdash \max\{P^k\}$ represents the remaining set after excluding the maximum value from $\{P^k\}$. Higher B_t^1 or B_t^1 values means more bias in textual review.

Measuring Dimensional Rating Bias. Based on the dimensional ratings for each review, we can measure the dimensional rating bias, which describes the extent to which the overall rating is driven by ratings from a few dimensions. For each dimensional rating, we take its absolute difference with the overall rating as DIF = (dif_1, dif_2, ..., dif_K). Hence, a biased review will have skewed distribution on DIF. We define the following measurement to quantify the dimensional rating bias:

$$B_r^1 = (mean(DIF \vdash \min\{DIF\}) - \min\{DIF\}) \Big/ mean(DIF \vdash \min\{DIF\}) \text{ or}$$

$$B_r^2 = (mean(DIF \vdash \min\_2\{DIF\}) - \text{mean}\{\min\_2\{DIF\}\}) \Big/ mean(DIF \vdash \min\_2\{DIF\}), \quad (3)$$

where min(DIF) denotes the minimum value in DIF and min_2 (DIF) denotes the first two minimum values in DIF. A bigger B_r^2 or B_r^1 value indicates more bias in overall rating. Note that if we have observed dimensional ratings in MS systems. And we may directly use them to compute DIF and the dimensional rating bias measurements.

De-biasing Dimensional Ratings. The goal of de-biasing is to correct the overall rating for a biased review. The overall rating could be modeled as a weighted combination of multiple dimension ratings (Wang et al. 2010) $r_i = \sum_{k=1}^{K} \beta^k Pol_i^k$, s.t. $\sum_{k=1}^{K} \beta^k = 1$, where β^k is the weight of dimension rating Pol_i^k for review d_i. The idea of de-bias method is to learn the weight with unbiased overall ratings and use the learned weights to estimate or modify the biased overall ratings. The rating bias in reviews with neutral overall ratings is expected to be less than that in reviews with extreme overall ratings as shown on the above. Thus, we propose to fit the regression model with neutral overall ratings (i.e., 3) and dimension ratings and obtain the weights $\{\beta^k\}$. Then we use the learned weights to estimate individual overall rating with extreme values (i.e., 1, 2, 4 and 5) based on the learned dimension ratings. Finally we get the modified average overall ratings by averaging all estimated extreme overall ratings and original neutral overall ratings for each product.

Preliminary Evaluation

Data

In order to evaluate our method in both SR and MR systems, we collect two hotel review datasets from Expedia.com and TripAdvisor.com, which is corresponding to SR and MR system respectively. Both Expedia and TripAdvisor are one of the most popular online travel agencies where users can book travel-related reservations (e.g., hotel and flight) and write reviews about their experience. Specifically, Expedia allows consumers to provide an overall numerical rating and a textual review. TripAdvisor allows consumers to give an overall numerical rating, dimensional numerical ratings (e.g., locations, service, and rooms for a hotel) and a textual review. It is worth noting that TripAdvisor used to be an SR system but changed to an MR system in 2009. To make the comparison between two systems fair, we focus on reviews for the same group of 50 hotels New York City (NYC). The specific six hotel dimensions in TripAdvisor.com are Value, Rooms, Location, Cleanliness, Service, and Sleep Quality. In both TripAdvisor.com and Expedia.com, consumers are allowed to provide numerical ratings of a 5-point scale. For each hotel, we collected all reviews/ratings available at both sites. In total, we have 71,438 Expedia reviews/ratings and 99,653 TripAdvisor reviews/ratings.

Preliminary results

The Dimensional Preference Bias. We apply our EM algorithm to both Expedia (SR system) and TripAdvisor (MR system) data sets to calculate the dimensional probability for each textual review, and then compute two types of dimensional preference biases according to equation 2. The bias statistics are in Figure 3. First, the average of bias measurement (in terms of both B_t^1 and B_t^2) in Expedia is higher than that in TripAdvisor data, which suggests that MR systems (e.g., TripAdvisor.com) could effectively mitigate the review bias. The underlying reason may be that the multi-dimensional rating mechanism could possibly remind consumers to comment on more dimensions, rather than on one or two most dimensions that impress consumers most. Also, the dimensional preference bias in reviews with extreme overall ratings (e.g., 1 out of 5, 5 out of 5) is much more than that that in reviews with neutral overall ratings (i.e., 3 out of 5) in both Expedia and TripAdvisor data. In order to rule out selection bias, in future study, we need to do rigorous statistical tests to controlling additional variables.

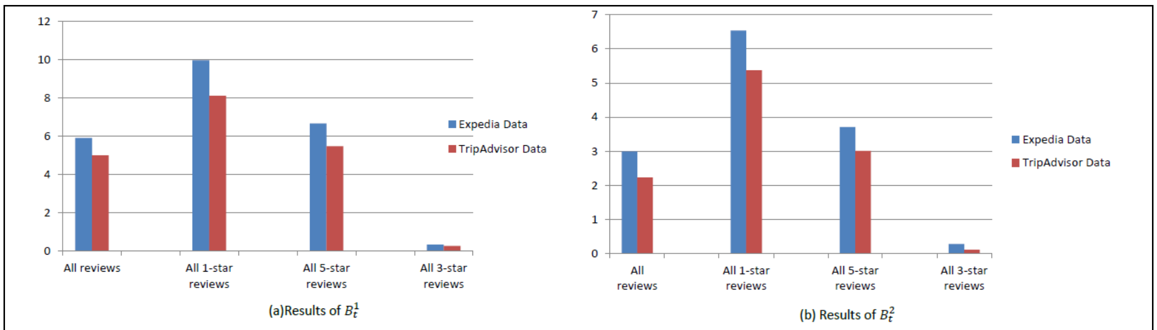

Figure 3. Results of the Dimensional Preference Bias

The Dimensional Rating Bias. We calculate the dimensional rating bias for both data sets based on the defined measurements in equation 3. The summary statistics are in Figure 4. As we can see, the average dimensional rating bias in Expedia data set is higher than that in TripAdvisor data set. This indicates that consumers are more likely to provide a biased overall rating in SR systems than in MR systems. In fact, in MR systems, when consumers write their reviews/ratings, systems will remind them to consider multiple dimensions of a product. Such a mechanism could potentially encourage users to accommodate ratings in all possible dimensions into their overall ratings. Also we can find that larger bias exists in reviews with extreme overall ratings than in those with neutral overall ratings.

The De-biasing Results. We show the original overall ratings and de-biased ones for randomly selected two hotels (i.e., Hotel 1 and Hotel 2), and the average absolute difference between the original overall ratings and de-biased ones over all hotels in Table 1, where Abs denotes the absolute value.

Expected Contribution and Future Work

Motivated by the Expectation-Confirmation Theory and Balance Theory, we identified two types of dimensional bias: dimensional preference bias and dimensional rating bias. We then proposed a novel approach to quantify these biases from textual reviews. Furthermore, we proposed a de-bias method for dimensional rating bias mitigation. Because our entire solution uses textual reviews and is independent of numerical ratings, it is applicable to both SR and MR systems.

Figure 4. Results of the Dimensional Rating Bias

Table 1. The De-bias Results

Website	Measures	Hotel1	Hotel2	All Hotels
Expedia (SR)	Original Rating	3.9	5	
	De-Biased Rating	4.5	4.6	
	Abs(Original-De-biased)	0.6	0.4	0.371
TripAdvisor (MR)	Original Rating	3.7	5	
	De-Biased Rating	4.1	4.8	
	Abs(Original-De-biased)	0.4	0.2	0.219

Our research contributes to the IS and marketing literature by proposing two quantitative dimensional bias measures. Although a handful of prior studies (e.g. Liu et al. 2014) have mentioned the importance dimensional bias, to the best of our knowledge, no studies have proposed a quantitative bias measure from textual reviews. It is probably due to the limited information provided by numerical ratings and complex knowledge required for converting textual reviews into numerical data. In this study, we demonstrate that customized data mining solution could close this gap. These two quantitative measures could spawn a stream of empirical research. For example, one could empirically test (1) whether these dimensional bias have a positive or negative effect on information sharing; (2) whether these bias are larger in SR than MR systems.

Furthermore, our research has a significant managerial impact. One interesting question that practitioners have been seeking an answer to is what reviews would solicit the highest product conversion. Prior studies have shed some light on the economic value of volume, valence, readability, and subjectivity of the reviews (Dellarocas et al. 2007, Duan et al. 2008a, Duan et al. 2008b, Liu 2006, Godes and Mayzlin 2004, Ghose et al. 2011), but not many of them have explored the economic value of the dimensional bias in the textual reviews. Moreover, managers could directly employ our bias measure to evaluate the quality of their review systems to support their business decision-making. Finally, business managers could utilize the proposed de-bias method to mitigate the risk introduced by dimensional bias.

We will complete this work by conducting a comprehensive and rigorous evaluation to demonstrate the reliability and validity of our measures. We will first conduct a rigorous offline evaluation by reporting metrics on an out-of-sample dataset. To make sure that our measure performs consistently across different domains, we will apply our method to a different domain, such as online retailer reviews. Additionally, we will conduct a user study to validate our measurements. Finally, as a robustness check, we will extend the Liu et al. (2014) study to demonstrate the impact of SR and MR systems on the dimensional bias extracted from textual reviews.

References

Adomavicius G, Tuzhilin A, Zheng R (2011). REQUEST: A query language for customizing recommendation, ISR 22(1): 99-117

Adomavicius, G., R. Sankaranarayanan, S. Sen, A. Tuzhilin. Incorporating Contextual Information in Recommender Systems Using a Multidimensional Approach. ACM Transactions on Information Systems, 23(1):103-145, January 2005.

Adomavicius, G., A. Tuzhilin. Multidimensional Recommender Systems: A Data Warehousing Approach. In L. Fiege, G. Mühl, U. Wilhelm (Eds.), Electronic Commerce: Second International Workshop (WELCOM 2001), Lecture Notes in Computer Science, vol. 2232, pp. 180-192, 2001.

Anderson, E. W., and Sullivan, M. W. 1993 "The Antecedents and Consequences of Consumer Satisfaction for Firms," Management Science (12:2), pp. 125-143.

Dellarocas, C., and C. Wood. 2007. "The Sound of Silence in Online Feedback:Estimating Trading Risks in the Presence of Reporting Bias," *Management Science, 54(3)*, pp. 460–476.

Dellarocas C, Zhang X, Awad NF (2007) Exploring the value of online product reviews in forecasting sales: the case of motion pictures. *Journal of Interactive Marketing* 21(4): 23-45.

Dimensional Research (2013). Customer service and business results: a survey of consumer service from mid-size companies

Duan W, Gu B, Whinston AB (2008a) Do online reviews matters? – An empirical investigation of panel data. Decision Support Systems 45(4): 1007-1016

Duan W, Gu B, Whinston AB (2008b) The dynamics of online word-of-mouth and product sales- An empirical investigation of the movie industry. J. Retailing 84(2): 233-242

Ghose A, and Ipeirotis P (2011) Estimating the helpfulness and economic impact of product reviews: ming taxt and reviewer characteristics, IEEE TKDE 23(10): 1498-1512.

Ghose, A., Ipeirotis P.G., and Li, B. (2012). "Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowd sourced content," Marketing Science (31:3), pp. 493-520.

Godes, D. and Silva, J. C. 2012. "Sequential and Temporal Dynamics of Online Opinion," *Marketing Science* (31:3), pp. 448-473.

Godes D, Mayzlin D (2004) Using online conversations to study word-of-mouth communication. *Marketing Science* 23(4): 545-560.

Heider, F. (1946). Attitudes and Cognitive Organization. Journal of Psychology, 21, 107–112.

Heider, F. (1958). The Psychology of Interpersonal Relations. New York: Wiley.

Hennig Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word of mouth via consumer opinion platforms: what motivates consumers to articulate themselves on the internet?. Journal of interactive marketing, 18(1), 38-52.

Hu, M., and Liu, B (2004). "Mining and summarizing customer reviews." *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.*

Krishnan, S., Patel, J., Franklin, M., Goldberg, K. 2014 "Social influence bias in recommender systems: a methodology for learning, analyzing, and mitigating bias in ratings," *Proceedings of the 8th ACM Conference on Recommender systems*, pp.137-144.

Li, X., and L. Hitt. 2008. "Self-Selection and Information Role of Online Product Reviews," *Information Systems Research, 19(4)*, pp. 456–474.

Liu, Y., Chen, P., and Hong, Y. 2014. "Matching Consumer Preferences with Product Attributes: The Value of Multi-dimensional Online Word of Mouth Systems," *Proceedings of ICIS.*

Liu, Y., (2006) Word of mouth for movies: Its dynamics and impact on box office receipts. *Journal of Marketing* 70(3):74-89

Moe, W. W., and Schweidel, D. A. 2012. "Online product opinions: Incidence, evaluation, and evolution," *Marketing Science* (31:3), pp. 372-386.

Moghaddam, S., and Ester, M. (2011). "Ilda: interdependent lda model for learning latent aspects and their ratings from online product reviews." *Proceedings of SIGIR.*

Moghaddam, S., and Ester, M. (2012). "On the design of lda models for aspectbased opinion mining." *Proceedings of CIKM.*

Muchnik, L., Aral, S., Taylor, S. J. 2013 "Social influence bias: a randomized experiment", *Science* (341), pp. 647-651.

Newcomb, T.M. (1953). An Approach to the Study of Communicative Acts. Psychological Review, 60, 393– 404.

NLTK. 2015. https://github.com/nltk/nltk/wiki/Sentiment-Analysis

Oliver, R. L. 1980. "A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions," Journal of Marketing Research (17:11), pp. 460-469.

Titov, I., and McDonald, R. (2008) "A joint model of text and aspect ratings for sentiment summarization." *Proceedings of ACL*.

UIC. 2015. http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html

Wang, H., Lu, Y., Zhai, C. (2010), "Latent aspect rating analysis on review text data: a rating regression approach", *ACM SIGKDD*, pp. 783-792.

Wang, H., Lu, Y., Zhai, C. (2011), "Latent aspect rating analysis without aspect keyword supervision." *ACM SIGKDD*.

Wang, T., Wang, D., Wang, F. 2014 "Quantifying herding effects in crowd wisdom", *ACM SIGKDD*, pp. 1087-1096.

Zajonc, R.B. (1971). The Concepts of Balance, Congruity, and Dissonance. In P. Suedfeld (Ed.), Attitude Change (pp. 63– 85). Chicago: Aldine-Atherton.